Code Generation: Results Achieved with Chain-of-Thought prompting Considering Only Methods with No External Dependencies
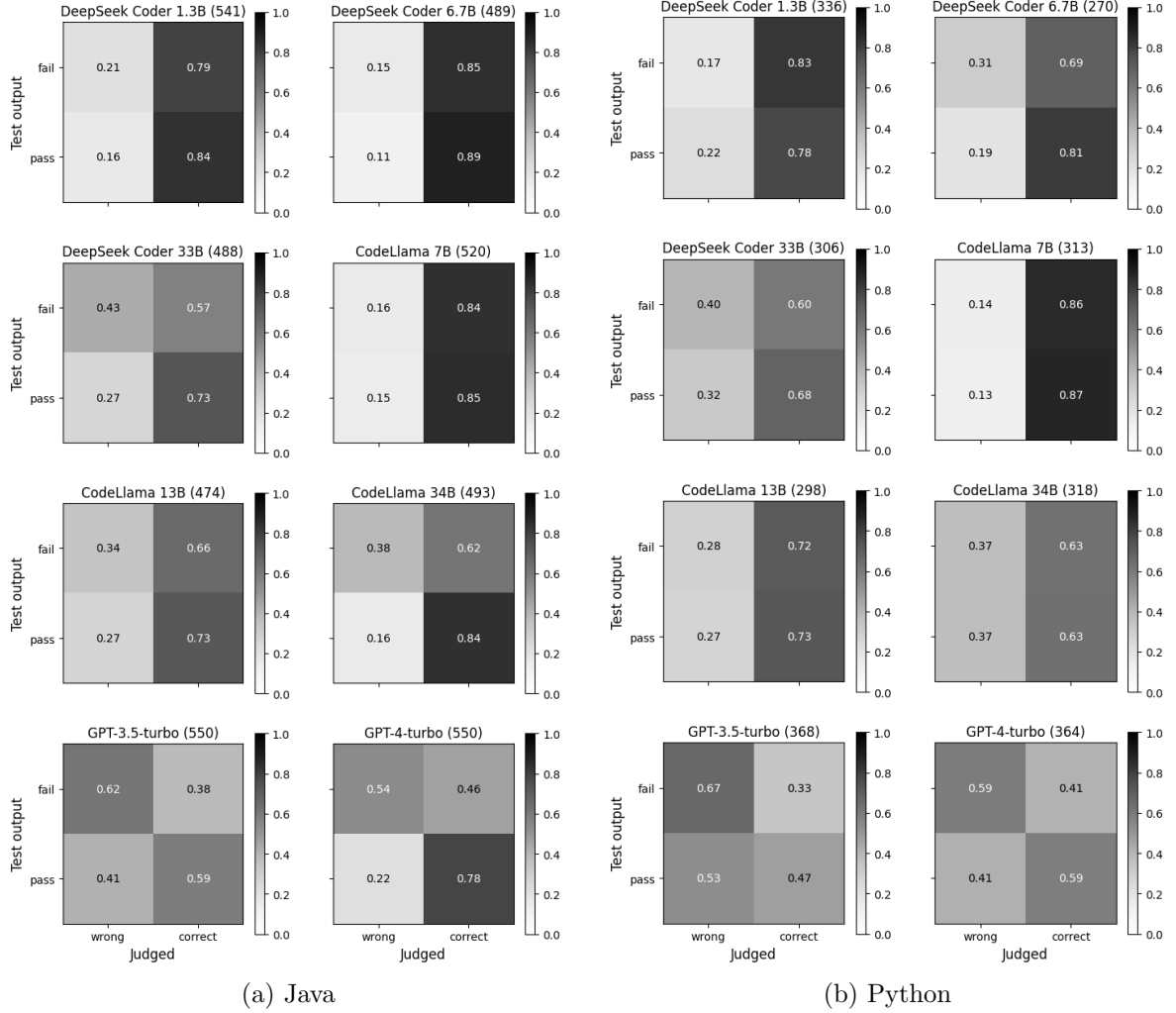

(a) Java

(b) Python

Figure 1: Code Generation: Confusion matrices for LLM's judgment.

	DSC 1.3B	DSC 6.7B	DSC 33B	CL 7B	CL 13B	CL 34B	GPT 3.5	GPT 4
Java	0.05	0.05	0.17	0.01	0.08	0.24	0.20	0.33
Python	-0.04	0.11	0.07	0.01	0.01	0.00	0.14	0.17

Table 1: Code Generation (only methods with no external dependencies): Kappa score.